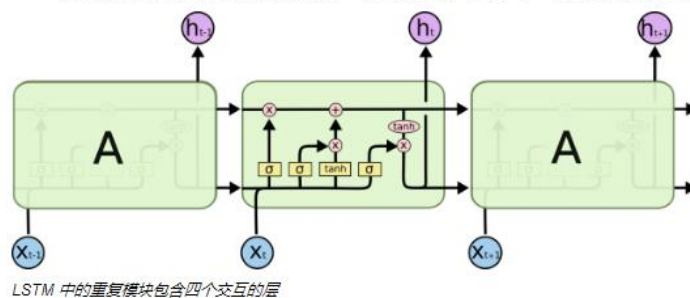
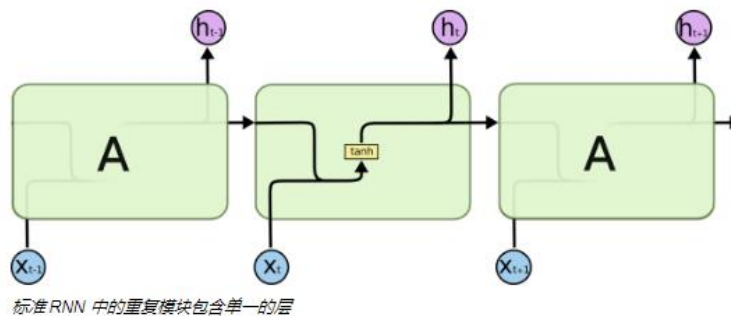


1. 深度学习内容的学习

我们的研究方向做以可视化作为特色的数据分析，重点为数据分析，现在最火的数据分析方法就是深度学习，因此对深度学习进行了实验。

- 学习 Python，搭建了 python+pycharm+tensorflow 的深度学习框架
- 了解深度学习的基础，从 CNN、RNN、LSTM 等框架。CNN（卷积神经网络）侧重点在空间方面，基本用在图像处理等多维数据的情况下。RNN（循环神经网络）侧重点在序列方面，多用于时间周期的预测上。RNN 只能参考短期记忆而不能利用长期序列的信息，因此有了 LSTM，在 RNN 的基础上完善了细胞结构，加入了门控制选项，可以将长期记忆特征加入的网络中来。



2. 工作思路

采用 Python 构建深度学习框架，并制定出交通事故合理的深度学习模型，完成交通事故预处理，作出基本预测。主要考虑一下几个问题：

- 针对于事故多发路段的交通事故数据如何转化为深度学习模型的输入数据
- 深度学习的网络如何设计深度有几层，宽度有几个
- 是针对一个多发路段构建一个深度学习模型，还是针对所有事故多发路段构建统一的深度学习模型

3. 最新想法（事故多发路段相似性判断）：

若在事故多发路段预测模型中，针对每一个事故多发路段都构建结构一样但参数不同的事故多发路段，那么可否根据深度学习模型中不同的参数来对事故多发路段完成相似性比对。

事故多发路段相似性比对是个比较复杂的问题，因为影响交通事故发生的因素比较复杂，且是由多方面的原因共同作用造成的。在事故多发路段相似性比对时，并没有明确的参考依据。但现在，假设我们采用统一的深度模型进行事故预测，对每一个事故多发路段都独立的进行训练，则每一个事故多发路段训练出来的参数都是不一样的。参数不一样的原因应该是交通事故对各个影响因素的敏感性造成的，则可以以深度学习模型的参数为依据，对事故多发路段进行相似性比对，对事故多发路段进行分类，且每类事故多发路段的诱发因素应该相似，则治理措施也应该相似。（这个问题的可行性还需要探讨）。

另一种事故多发路段：采用格兰杰因果分析实验进行判断：

（在时间序列情形下，两个经济变量 X 、 Y 之间的格兰杰因果关系定义为：若在包含了变量 X 、 Y 的过去信息的条件下，对变量 Y 的预测效果要优于只单独由 Y 的过去信息对 Y 进行的预测效果，即变量 X 有助于解释变量 Y 的将来变化，则认为变量 X 是引致变量 Y 的格兰杰原因。）

则可以根据控制参与深度学习模型预测的数据类型得到的预测准确度对原因的影响因素进行排序，但该方法没有很强烈的因果性。

4. 文献阅读：

Zhang Z, He Q, Gao J, et al. A deep learning approach for detecting traffic accidents from social media data[J]. *Transportation Research Part C: Emerging Technologies*, 2018, 86:580-596.

（从社交数据出发，通过深度学习的方法来识别其中的交通事故。）本文采用深度学习的方法从社交媒体数据中检测交通事故。文章调查了两个大城市北弗吉尼亚和纽约市，1年内超过300万的tweet内容。结果表明，成对的tokens可以捕获与事故相关的tweet中固有的关联规则，进一步提高交通事故检测的准确性。文章采用了两个深度学习网络：Deep Belief Network (DBN)和Long Short-Term Memory (LSTM)。DBN的分类结果比LSTM、SVM、sLDA方法都要好。通过对比，提出了利用Twitter检测交通事故的几个重要问题，包括位置和时间偏差，以及有影响力用户的特点和标签。

G. Zhang, K. K. W. Yau, and G. Chen, "Risk factors associated with traffic violations and accident severity in China." *Accident; analysis and prevention*, vol. 59, pp. 18–25, Oct. 2013.

（详细的说明了交通事故的影响因素，重点家中了交通违章对事故的影响，违章数据在目前的交通事故预测分析中采用的还比较好，而违章数据应该也是我们容易拿到的数据）在本研究中，分析了2006-2010年广东省的交通事故数据。我们特别关注两个结果指标：交通违规和事故严重程度。考虑了人、车辆、道路和环境的危险因素。首先，研究结果确立了交通违法行为作为威胁道路安全的主要风险之一的作用。确定了与交通违章和事故严重程度相关的特定风险因素

Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial temporal correlation in a hybrid deep learning framework," *arXiv preprint arXiv:1612.01022*, 2016

深度学习方法在人工智能领域已经达到了一个名人的地位，它的成功主要依赖于卷积网络(CNN)和递归网络。利用图像和视频的基本空间属性，CNN在视觉任务上总是取得优势表现。而递归网络(RNN)尤其是长短时记忆方法(LSTM)能够成功地表征时间相关性，从而表现出对时间序列任务的优越能力。交通流数据具有丰富的时空特征。然而，CNN和LSTM方法在交通流中的应用有限。在本文中，我们提出了一种结合CNN和LSTM的新型深度架构来预测未来的交通流(CLTFP)。（将CNN与LSTM相结合构建出新的深度学习网络，来进行短期交通流预测）利用一维CNN获取交通流的空间特征，利用两个LSTMs挖掘交通流的短期变异性周期。基于这些有意义的特征，进行特征级融合，实现短期交通流预测。将该方法与开放数据集上其他常用的预测方法进行了比较。实验结果表明，该方法在交通流预测中具有较大的优势。此外，从格兰杰因果关系的角度分析了所提出的CLTFP，发现并讨论了交通流和CLTFP的几个有趣的性质。

F. Toque, E. C'ome, M. K. El Mahrsi, and L. Oukhellou, "Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 1071–1076.

（采用深度学习方法，进行OD数据的预测）在使用智能卡数据分析城市流动方面进行了大量的研究。这些研究的主要目的是识别反复出现的乘客习惯、揭示出行模式、重建和预测客流等。对公共交通部门和运营商来说，预测出行需求都是一个核心问题。这是有效分配和优化现有资源的第一步。摘要提出了一种利用长短时记忆

递归神经网络预测地铁网络动态始发目的地矩阵的新方法。通过对法国雷恩地铁公司公共交通网络的实测智能卡数据集,与传统的日历法、向量自回归法等方法进行了比较。获得的结果表明,可靠的短期预报(在 15 分钟的时间范围内)的 OD 对提出的方法可以实现。实验了考虑附近交通系统(本例中为公交车)OD 矩阵的附加数据对预测精度的影响。

S.-h. Park, S.-m. Kim, and Y.-g. Ha, “Highway traffic accident prediction using vds big data analysis,” *The Journal of Supercomputing*, vol. 72, no. 7, pp. 2815–2831, 2016

交通事故的预测可以通过分类分析来实现,分类分析是一种数据挖掘过程,需要足够的数据来建立一个学习模型。然而,建立这样一个预测系统涉及几个问题。交通事故预测需要大量的硬件资源来收集和分析交通数据。此外,与交通事故相关的数据量小于与交通事故无关的数据量;这两类(预测类和其他类)数据的数量不同,因此是**不平衡的**。**(数据的不平衡性即交通事故发生的情况比正常情况要少很多,导致训练出来的结果不准确,这个问题可以作为下一步研究的内容。)**本文的目的是建立一个能够解决这些问题的预测模型。本文建议使用 Hadoop 框架对大流量数据进行高效的处理和分析,并提出一种解决数据不平衡问题的采样方法。在此基础上,预测系统首先对大流量数据进行预处理,并对其进行分析,为学习系统创建数据。使用抽样方法纠正创建数据的不平衡。为了提高预测精度,将校正后的数据分为若干组,并对其进行分类分析。

5. 论文整理

面向事故多发路段的交通事故风险预测系统

1. 前言

近年来,随着经济的发展,城市车有量越来越多,引起交通拥堵、环境污染及交通事故等一系列问题。世界卫生组织在 2015 年发表《全球道路安全状况报告》中指出每年约有 125 万人死于交通事故。交通事故已成为现今社会急需解决的问题。

通过交通大数据和深度学习,实时的交通流量预测使得人们可以通过选择不那么拥堵的路线来避免交通拥堵。通过交通大数据和深度学习也可能为预测或降低交通事故风险提供一个有前途的解决方案。从历史交通事故数据通过大数据分析和深度学习,识别交通事故中的规律并加以治理是现在交通事故领域的主要研究问题。

建立有效的交通事故风险预测系统是交通事故预防中的一项重要任务。如果某一区域的交通事故风险是可以预测的,我们可以将这一信息传播给附近的司机,提醒他们或者让他们选择一条不那么危险的道路;交警部门也可以在该区域有针对性的安排巡逻人员。然而交通事故风险的准确预测是非常困难的,许多因素都会影响交通事故。例如,不同地区的交通事故率有很大的差异。此外恶劣的天气条件例如雪或者大雾,也会降低道路的能见度和交通能力,从而增加交通事故的发生几率,一天中不同时间的交通事故发生率也不同。虽然许多研究人员已经把注意力集中在与交通事故相关的关键因素的识别上【1】,交通事故风险动态有效预测仍然是一个具有挑战性的问题。

随着深度学习的发展,基于深度学习和大数据的方法在交通相关问题上取得了良好的效果,例如交通流量预测【2】、到达时间估计【3】、起始结束点分析【4】。对于基于深度学习的交通事故风险预测,Chen 等人通过 Stack 去噪自动编码器来提取的人的移动特征【5】,从而推断日本境内交通事故风险。Ren 等人综合考虑交通事故的周期性模式和空间分布模式,【6】并结合其他重要因素例如:天气状况、空气质量等,采用了 LSTM 深度学习模型,构建了北京市的交通事故风险预测系统。但他们都是将整个城市区域划分成正方形的小块,针对每一个小块进行事故风险预测。每一个小块包含大量的非道路部分,而在这些区域通常不会有交通事故的发生。因此,本文以在识别事故多发路段的基础上,面向事故多发路段构建了交通事故预测模型。

2. 相关文献

2.1 事故多发路段识别文献介绍

交通事故、伤亡事故是可以预测和预防的，但事故易发地点的确定是分析事故原因并采取有效对策的前提。事故多发地区是一个定义不严格的术语，指的是容易发生比平均数量更多事故的黑点和/或区域。事故多发路段的识别基本有以下几种方法。

事故频率法【7】是一种以事故数量为指标的常规统计方法。在确定的时间和地点。如果事故数量大于阈值，则将其确定为事故易发位置。同样，事故率法【8】使用事故率(单位道路长度事故数、单位车辆公里事故数或单位交通流量事故数)作为指示指标。多学科方法【9】以事故数量为水平坐标轴，事故数量为纵向坐标轴，计算事故频率和发生率的分布。整个坐标可以分为四个区域:高事故率和高事故率区域，高事故率和低事故率区域，低事故率和高事故率区域，事故率和低事故率区域。分为第一个区域的路段可以划分为事故多发区，分为最后一个区域的路段可以划分为相对安全的区域。其他两个部分则有待进一步分析。事故累积频率曲线法【10】计算事故累积频率曲线，以单位长度事故数为横坐标，大于事故发生频率的累积频率为纵坐标。通常情况下，较小的事故分散在较大比例的路段，而较大的事故集中在较小比例的路段。事故发生频率越高，累计发生频率越低。因此，如果累积频率较小，而事故频率较高，则可以区分出事故多发区域。在上述方法中，道路单位长度是划分路段计算事故发生频率和/或发生率的重要参数，导致事故易发地段发生泄漏。为了消除这个缺点，应该引入一个比单位长度小得多的步长来对附近区域进行详尽的搜索。但是，它不能从根本上解决这个问题，而且它会极大地增加计算负担，尤其是在步长很小的情况下。统计方法将道路划分为单位长度的路段，存在着道路渗漏的缺点。基于聚类【11】的方法不存在这个问题，不需要道路分割。基于密度的空间聚类噪声应用(DBSCAN)【12】是目前最流行的基于密度的聚类算法之一。

2.2 实时交通事故风险预测

随着机器学习的发展，许多研究者开始关注交通事故的实时预测。Lv 选择基于欧氏度量的特征变量，利用 k 近邻法预测交通事故【13】。Park 收集了首尔高速公路的大交通事故数据，建立了基于 k-means 聚类分析和 logistic 回归的预测工作流【14】。最近，Chen 利用日本的人类移动数据，构建了一个 Stack 去噪自动编码器，来推断实时交通风险【5】。这些作品的一个局限性是没有将交通流量、天气状况、空气质量等几个重要因素纳入模型。如果没有这些信息，模型的预测能力可能会被削弱。

2.3 深度学习

深度学习的成功证明了它在发现高维数据中复杂结构方面的能力。它作为图像识别、语音识别、自然语言理解等领域的最新技术而得到了广泛的应用。对于智能交通系统的研究，很多研究都是基于深度学习的交通流预测。在较长的时间尺度上，一些研究试图预测大规模交通网络【15】的拥堵演化。另一个有趣的应用是利用深度强化学习来控制交通信号灯的控制。

3. 数据介绍与功能分析

3.1 数据介绍

3.1.1 交通事故数据、天气数据、交通流数据、节假日数据

3.1.2 系统功能分析: 预测事故多发路段实时发生交通事故的风险(采用深度学习的方法)

4. 整体思路介绍

(想预测事故多发路段实时发生交通事故的风险，首先必须要有事故多发路段，因此本系统的第一步仍是，识别事故多发路段)

(第二部，基于事故多发路段，引入事故多发路段的道路信息，加入交通流量与天气数

据，构建深度学习预测模型，完成事故多发路段事故发生风险的预测）

4.1 系统的整体结构介绍

识别事故多发路段

由事故多发路段作为识别对象，识别事故多发路段发生交通事故的风险

优点：大部分交通事故发生在事故多发路段处，从事故多发路段出发能保证预测的更加合理；事故多发路段定义在道路上，可以引入道路的基本信息。

4.2 事故多发路段识别

（识别事故多发路段的方法采用我们从前设计的道路危险度的概念。在之前工作的基础上，设计自动阈值识别算法，自动识别合适的阈值）

4.2.1 道路危险度概念

4.2.2 事故多发路段识别步骤

- 1) 计算道路上某点单起交通事故对的道路危险度的影响
- 2) 计算道路上某点一段时间内的道路危险度
- 3) 计算道路上所有点的道路危险度
- 4) 将道路上道路危险度大于阈值的点聚合成段即位事故多发路段。

4.2.3 阈值自动确定算法

（阈值自动识别的算法，参考该文献 Qiu C , Xu H , Bao Y . Modified-DBSCAN Clustering for Identifying Traffic Accident Prone Locations[M]//Intelligent Data Engineering and Automated Learning – IDEAL 2016. Springer International Publishing, 2016.）

该文献采用聚类算法，提出了一种自动确定参数的方法，我们可以引入过来，依据识别的结果自动确定阈值参数。

4.3 基于深度学习的预测模型

采用深度预训练进行实时交通事故预测的论文有 2 篇：

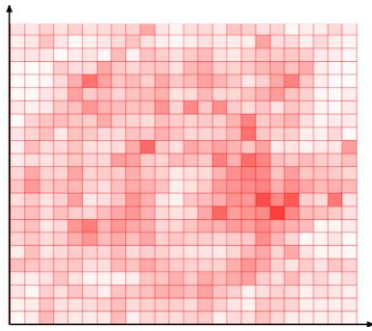
[1]Q. Chen, X. Song, H. Yamada, and R. Shibasaki, “Learning deep representation from big and heterogeneous data for traffic accident inference,” Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16) 2016.

[2] Ren H , Song Y , Wang J , et al. A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction[J]. 2017.

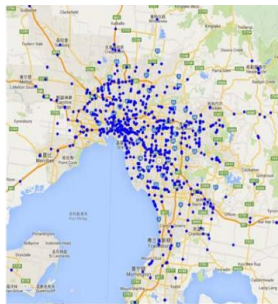
文献【1】仅采用人口移动模型，采用 Stack denoise Autoencoder 构建了深度学习网络，预测日本交通事故发生风险的发生

文献【2】引入了更多的基础数据，如天气，节假日，交通流量信息，采用 LSTM 模型构建了深度学习网络。

2 篇论文都是直接把整个研究区域划分成 1km*1km 的平方块，预测每一个平方快的结果。



对区域进行分块处理，考虑了大量的无关数据，交通事故预测，研究的重点应该是事故多发区域，因此本文提出改进，不在对整个研究区域进行分块，而是将事故黑点或者事故多发路段作为基本研究单元。



事故黑点



事故多发路段

（采用深度学习的框架，进行预测）

5. 验证分析

基于合适的可视化模型

6. 下一步工作

通过控制输入变量，通过预测结果的准确性，判断各个事故多发路段的原因重要性排序

参考文献

- 【1】 G. Zhang, K. K. W. Yau, and G. Chen, "Risk factors associated with traffic violations and accident severity in China." *Accident; analysis and prevention*, vol. 59, pp. 18–25, Oct. 2013
- 【2】 Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatialtemporal correlation in a hybrid deep learning framework," *arXiv preprint arXiv:1612.01022*, 2016.
- 【3】 C. Bai, Z. Peng, Q. Lu, and J. Sun, "Dynamic bus travel time prediction models on road with multiple bus routes," *Computational Intelligence and Neuroscience*, vol. 2015, pp. 432 389–432 389, 2015
- 【4】 F. Toque, E. C'ome, M. K. El Mahrsi, and L. Oukhellou, "Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 1071–1076
- 【5】 Q. Chen, X. Song, H. Yamada, and R. Shibasaki, "Learning deep representation from big and heterogeneous data for traffic accident inference," 2016.
- 【6】 Ren H , Song Y , Wang J , et al. A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction[J]. 2017.
- 【7】 Oppe, S.: Development of traffic and traffic safety: global trends and incidental fluctuations. *Accid. Anal. Prev.* **23**(5), 413–422 (1991)
- 【8】 Wright, C.C., Abbess, C.R., Jarrett, D.F.: Estimating the regression-to-mean effect associated with road accident black spot treatment: towards a more realistic approach. *Accid. Anal. Prev.* **20**(3), 199–214 (1988)

- 【9】 Larsen, L.: Methods of multidisciplinary in-depth analyses of road traffic accidents. *J. Hazard. Materials* **111**(1–3), 115–122 (2004)
- 【10】 Fang, S., Guo, Z., Yang, Z.: A new identification method for accident prone location on highway. *J. Traffic Transp. Eng.* **1**, 91–98 (2001)
- 【11】 Xie, L., Chaozhong, W., Lyn, N., Gao, Y.: Accident-prone section identification approach based on improved clustering algorithm. *J. Wuhan Univ. Technol. (Transp. Sci. Eng.)* (4), 904–908 (2014)
- 【12】 Ester, M., Kriegel, H.-P., Sander, J., Xiaowei, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 226–231 (1996)
- 【13】 Y. Lv, S. Tang, and H. Zhao, “Real-time highway traffic accident prediction based on the k-nearest neighbor method,” in *Measuring Technology and Mechatronics Automation, 2009. ICMTMA '09. International Conference on*, vol. 3. IEEE, 2009, pp. 547–550.
- 【14】 S.-h. Park, S.-m. Kim, and Y.-g. Ha, “Highway traffic accident prediction using vds big data analysis,” *The Journal of Supercomputing*, vol. 72, no. 7, pp. 2815–2831, 2016.
- 【15】 X. Ma, H. Yu, Y. Wang, and Y. Wang, “Large-scale transportation network congestion evolution prediction using deep learning theory,” *PloS one*, vol. 10, no. 3, p. e0119044, 2015.